Vision Research 107 (2015) 101–112

Contents lists available at ScienceDirect

Vision Research

journal homepage: www.elsevier.com/locate/visres

The effective reference frame in perceptual judgments of motion direction

Mehmet N. Agaoglu [a,b,*], Michael H. Herzog [c], Haluk Öğmen [a,b]

[a] Department of Electrical and Computer Engineering, University of Houston, N308 Engineering Building 1, Houston, TX 77204-4005, USA
[b] Center for Neuro-Engineering and Cognitive Science, University of Houston, Houston, TX 77204-4005, USA
[c] Laboratory of Psychophysics, Brain Mind Institute, Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland

ARTICLE INFO

Article history:
Received 13 May 2014
Received in revised form 26 November 2014
Available online 20 December 2014

Keywords:
Motion perception
Reference frames
Eye movements

ABSTRACT

The retinotopic projection of stimulus motion depends both on the motion of the stimulus and the movements of the observer. In this study, we aimed to quantify the contributions of endogenous (retinotopic) and exogenous (spatiotopic and motion-based) reference frames on judgments of motion direction. We used a variant of the induced motion paradigm and we created different experimental conditions in which the predictions of each reference frame were different. Finally, assuming additive contributions from different reference frames, we used a linear model to account for the data. Our results suggest that the effective reference frame for motion perception emerges from an amalgamation of motion-based, retinotopic and spatiotopic reference frames. In determining the percept, the influence of relative motion, defined by a motion-based reference frame, dominates those of retinotopic and spatiotopic motions within a finite region. We interpret these findings within the context of the Reference Frame Metric Field (RFMF) theory, which states that local motion vectors might have perceptual reference-frame fields associated with them, and interactions between these fields determine the selection of the effective reference frame.

1. Introduction

When an object moves during steady fixation, its projection on the retina also moves at a speed proportional to its physical speed. The perceptual system readily interprets this retinal motion as the motion of an object in the environment. However, when the observer's eyes, head or body move, the retinal image motion does not directly correspond to a corresponding motion in the environment. In order to perceive veridically the motion of an object in the environment, the perceptual system needs to carry out coordinate transformations (Swanston, Wade, & Day, 1987; Wade & Swanston, 1987). In other words, the retinal motion due to self-motion or movement of the eyes need to be parsed out such that what is left directly corresponds to the motion of an object in the environment. Gibson argued that optic flow alone is sufficient to make the required transformations and to decompose retinal motion into self-motion and object motion relative to the scene (Gibson, 1979). Many psychophysical (e.g. Rushton, Bradshaw, & Warren, 2007; Warren & Rushton, 2009), neurophysiological (Duffy & Wurtz, 1991a; Duffy & Wurtz, 1991b), functional imaging (e.g. Morrone et al., 2000), and modeling (Furman & Gur, 2003; Pack, Grossberg, & Mingolla, 2001) studies supported his position. However, early studies of motion perception during smooth-pursuit eye movements showed that the coordinate transform from retinocentric reference frame to head-centric one is not perfect. A stationary object is perceived to be moving in the direction opposite to the direction of the ongoing pursuit eye-movement (Filehne illusion. Filehne, 1922; Freeman & Banks, 1998; Mack & Herman, 1972; Mack & Herman, 1973; Wertheim, 1987) and a moving object is perceived to be slower when it is tracked than when it is viewed during fixation (Aubert-Fleischl effect. Fleischl, 1882; Aubert, 1886; Freeman & Banks, 1998). The perceived direction and the extent of motion of an object that moves non-collinearly with the pursuit target significantly deviate from corresponding physical quantities (Becklen, Wallach, & Nitzberg, 1984; Festinger, Sedgwick, & Holtzman, 1976; Furman & Gur, 2005; Kano & Hayashi, 1981; Souman, Hooge, & Wertheim, 2005; Souman, Hooge, & Wertheim, 2006b). Assuming perfect retinal gains (i.e. the ratio of perceived and actual retinal motion extents or speeds is 1), these perceptual errors and illusions have been conventionally attributed to an under-registration of eye velocities. However, perceived retinal motion is strongly modulated by stimulus properties such as spatial frequency, dot density, contrast,

* Corresponding author at: Department of Electrical and Computer Engineering, University of Houston, N308 Engineering Building 1, Houston, TX 77204-4005, USA.

http://dx.doi.org/10.1016/j.visres.2014.12.009
0042-6989/

stimulus scale and chromatic content (see review by Nishida (2011)) and hence, errors in estimating retinal motion should also be considered in the computations of head-centric motions (Freeman & Banks, 1998).

Many models of motion perception during smooth pursuit have been proposed to quantify the degree to which this coordinate transformation is complete. In most of these models, the observer's head and body are assumed to be stationary with respect to the outside world, and the perceived head-centered motion is a combination of retinal motion and eye velocity estimates (Freeman, 2001; Freeman & Banks, 1998; Souman, Hooge, & Wertheim, 2006a; Swanston et al., 1987; Turano & Massof, 2001; Wertheim, 1994). Models with non-linear motion transducers have been shown to perform slightly better than those with linear motion estimators for both terms (Freeman, 2001; Turano & Massof, 2001). The estimated eye velocity in some of these models is a function of both retinal and extra-retinal signals, whereas retinal motion estimates depend only on stimulus parameters and retinal motion itself (Freeman & Banks, 1998; Turano & Massof, 2001; Wertheim, 1994). Several studies have concluded that perceived motion during pursuit also depends on stimulus parameters including size (Turano & Heidenreich, 1999), spatial frequency (Freeman & Banks, 1998; Wertheim, 1994), speed (Pola & Wyatt, 1989; Turano & Heidenreich, 1996), and presentation duration (Mack & Herman, 1978; Souman et al., 2005; Wertheim, 1987).

When there are two objects in the scene and one of them is tracked, the relative motion between the objects may become a major determinant of perceived motion. In some studies, this fact was overlooked and the failure to discriminate relative motion from retinal motion led some researchers to conclude that the perceptual system has very weak (i.e. gains < 0.1) or no information at all about the ongoing pursuit eye movement (Dodge, 1904; Festinger et al., 1976; Stoper, 1973). For instance, when a small dot is pursued in a dark room and the motion of another (moving or stationary) dot is judged, the retinal motion of the target dot and its relative motion with respect to the pursuit dot are almost identical (assuming perfect smooth pursuit). It is impossible to decouple contributions of the retinal and relative motions in these displays. In fact, Mack and Herman (1978) showed that the relative motion between the pursuit target and the background object is one of the main factors influencing perceived motion. The contribution of the relative motion between the pursuit target and the background has been noted in several studies (Baker & Braddick, 1982; Brenner & van den Berg, 1994; Freeman, Champion, Sumnall, & Snowden, 2009; Freeman, Champion, & Warren, 2010; Hisakata, Terao, & Murakami, 2013; Mack & Herman, 1978; Mateeff, Hohnsbein, & Ehrenstein, 1990; Snowden, 1992; Turano & Heidenreich, 1999; Wallach, 1959; Wallach, O'Leary, & McMahon, 1982). In these studies, qualitative descriptions of how and when relative motion between the pursuit target and the background affects perceived motion have been given. Baker and Braddick (1982) argued that, at slow speeds, relative motion determines percepts whereas at high speeds, absolute motion (i.e. motion with respect to a spatiotopic reference frame such as stimulus display) takes over. Mack and Herman (1978) concluded that object-relative motion is only effective when the object of interest is in close proximity of the pursuit target. Brenner and van den Berg (1994) reported that the perceived target velocity does not change as long as the relative motion of the pursuit target with respect to a textured background is kept fixed.

When there are multiple moving objects in the scene, a typical scenario in normal viewing conditions, relative motions of these objects can fully determine the perceived motion. Duncker (1929) used displays generated by point-lights attached to an otherwise invisible rotating and translating circular cardboard (Duncker, 1929, pg. 240). When a point-light is attached to the rim of the cardboard, observers perceive cycloidal motion of the light, which corresponds to its trajectory on the retina if the observer's eyes are stationary. Percepts do not change when the point-light is tracked. However, when another point-light is added to the hub of the wheel, the central light is perceived to be translating linearly, whereas the peripheral light is perceived as rotating around the central light, regardless of whether the central light is tracked or not. In the latter case, the retinal trajectory of the point-light at the rim is again a cycloid; but the percepts dominantly correspond to its relative motion with respect to the central light. Similar and more complex demonstrations of the superiority of relative motion were done Johansson (1950) and Johansson (1973). In line with this, it has been shown that the thresholds for detecting relative motion is much less than those for absolute motion (Snowden, 1992). Moreover, the movements of the eyes, head or body result in relative motions of objects at different depths in the environment. A complete theory of motion perception, therefore, must take into account the relative motion of objects with respect to each other. Wade and Swanston's quantitative model of motion perception (Wade & Swanston, 1987) explicitly includes a term for relative motion of objects with respect to each other. According to their model, the registered retinal motion undergoes a sequence of coordinate transforms to reach a geocentric representation. Estimated retinal motions are compensated for estimated eye movements at the orbital level, and the output of this process is combined with the "pattern-centric" signals (i.e. relative motion). Furthermore, they proposed that the two signals are not treated equally, but each has a weight. A similar approach was taken by Gogel (1977). He also argued that the relative motion has a greater weight compared to the other components (Gogel, 1977). Unfortunately, the weights of different terms have never been determined experimentally.

In contrast to the models of motion perception mentioned so far, we adopted a top-down approach and modeled the perceived motion as an interplay between various reference frames available to the perceptual system. By doing so, we remained agnostic as to how coordinate transforms outlined by previous models take place; instead, we sought to investigate how the perceptual system forms the "effective reference frame". Let's assume that the head is kept still and two objects are moving in the fronto-parallel plane at different velocities. The perceived motion of each object depends on its motion on the retina (i.e. retinocentric or retinotopic reference frame), its motion on the display (i.e. space-centric or spatiotopic reference frame), and its motion relative to the motion of the other object (i.e. object-based or motion-based reference frame). The proposed model is given by

$$\text{Perceived motion} = w_s(d,\varphi)P_s + w_r(d,\varphi)P_r + w_{mb}(d,\varphi)P_{mb} + c, \tag{1}$$

where P_s, P_r, and P_{mb} represent the motion signals on spatiotopic, retinotopic and motion-based reference frames, and w_s, w_r, and w_{mb} represent the weights of each reference frame, respectively. The constant term c in the model captures the response bias of observers. The response bias represents byproducts of decision processes. Each P value represents also the predicted perceived-motion from a given reference frame. For instance, if observers perceive the motion direction solely based on retinal motion, (i.e. $w_s = 0$, $w_{mb} = 0$, and $w_r = 1$), perceived motion would be equal to P_r. Note that each weight is modeled as a function of distance d between the two objects and some other potential factors φ (such as perceptual groupings, stimulus scale, attention, etc.).

Equation (1) contains four unknowns, namely the three weights and the constant term. In order to have a unique solution, at least four linearly independent equations (i.e. different combinations of P_s, P_r, and P_{mb} values) are needed. To this end, we designed four experimental conditions where observers judged whether or not a horizontally moving small (target) disk reversed its direction of motion (see methods for details). Our stimulus is depicted in Fig. 1A. One small (target) and one large (reference) disk moved horizontally. The large disk moved with a constant velocity while the velocity of the small disk changed sinusoidally during the mid-course of its trajectory (Fig. 1B). The average velocity of the small disk was equal to the velocity of the large disk so as to keep the average distance between the two disks constant. This constraint was introduced because the distance between objects is known to affect the extent to which relative motion is perceived (Gogel, 1974; Gogel & Koslow, 1972; Mack & Herman, 1978; Mateeff & Hohnsbein, 1989; Mori, 1979; Shum & Wolford, 1983). In several studies of relative motion perception, stimuli consisted of objects with different velocities. As a consequence, the distance between the objects changed during the stimulation period. This resulted in a confound between the distance effect and the relative-motion effect (Festinger et al., 1976; Mack & Herman, 1978; Stoper, 1973). This issue becomes even more severe when the two objects move non-collinearly. To avoid this confound, we kept the *average* distance between the two objects constant.

A variety of stimuli, pursuit conditions, and tasks have been used in different studies to investigate perceived motion during eye movements. It is known that stimulus parameters such as size, speed, task, and attention affect performance in speed judgments (Baker & Braddick, 1982; Freeman et al., 2009; Gogel & Sharkey, 1989; Mateeff et al., 1990). Therefore, it is difficult to compare and reconcile results of different studies. Here, we investigated various pursuit conditions with the same task, stimuli and procedures as described below.

The rationale of our experimental conditions was based on four considerations: First, we wanted to quantify the contributions of different reference frames in determining the effective reference frame for motion perception. In order to dissociate between retinotopic and spatiotopic reference frames, we included conditions with and without eye movements. Second, we wanted to capture the quantitative contributions of different reference frames into a simple mathematical model, as given in Eq. (1). Since the model in Eq. (1) has four free parameters, four linearly independent equations are needed to estimate these parameters. Hence, we designed four conditions as follows: In the Baseline condition, a target disk was viewed during steady fixation (Fig. 1A, Baseline). In the Fixation condition (Fig. 1A, Fixation), another moving disk was also presented while observers kept fixation at the center of the display. In the smooth pursuit (SP) Same and SP Opposite conditions (Fig. 1A, SP Same and SP Opposite), observers tracked with smooth pursuit eye-movements another disk moving either in the same or in the opposite direction of the target disk, respectively. Third, we also varied the distance between the horizontally moving disks to characterize the weights in Eq. (1) as a function of d to address previous accounts of distance dependent effects (Gogel, 1974; Mack & Herman, 1978). Finally, we wanted to test the predictive ability of

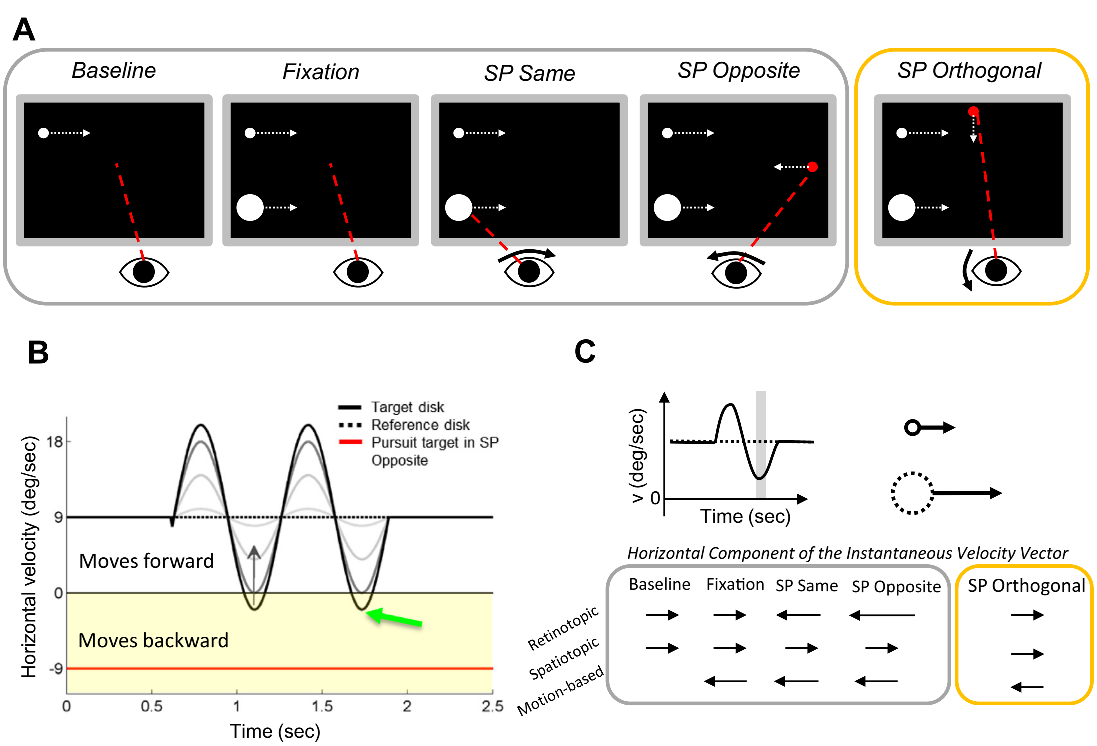

Fig. 1. Spatial and temporal characteristics of the stimuli. (A) Stimuli used in different experimental conditions. A small white disk (target) was shown on a dark background in all conditions. In all conditions, with the exception of the Baseline condition, another white disk (reference), larger in size, was also shown. The small red dot shown in the smooth pursuit (SP) opposite and SP Orthogonal conditions represents the pursuit target specific to these conditions. See text for detailed explanations. Results from the Baseline, Fixation, SP, and SP Opposite conditions were used to predict observers' percepts in the SP Orthogonal condition. (B) Horizontal velocity profiles of the target (black/gray curves) and reference (black dotted line) disks are shown as a function of time. The reference disk always moved with constant speed whereas the velocity of the target disk was modulated by a sine wave in the central half of the stimulus presentation. The minimum speed of the target disk was the dependent variable in a staircase algorithm and was modulated by varying the amplitude of sine wave. If the amplitude of the sine wave exceeds the average speed (i.e. 9 deg/s), the minimum velocity falls below 0 deg/s and the target dot moves backward according to a spatiotopic reference frame, for a short temporal interval. The task of the observers was to report whether the target dot moved backward at any point in time during its motion on the display. The red horizontal line represents the velocity of the smooth pursuit target in SP Opposite condition. The smooth pursuit target in the SP Orthogonal condition moved in the vertical dimension. C. Instantaneous horizontal velocity vectors (bottom) of the target disk on retinotopic, spatiotopic and motion-based reference frames shown for an example motion profile (top-left). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

the quantitative model. For this purpose, we have included a fifth condition, SP Orthogonal (Fig. 1A, SP Orthogonal), where observers pursued an additional vertically moving object with the goal of testing whether the weights obtained from the results of the first four conditions can be generalized to this new condition.

Fig. 1B illustrates the velocity profiles used in the experiments and in Fig. 1C we provide a specific example of velocity profiles for the target and reference disks, along with their horizontal instantaneous motion vectors on each reference frame when the target disk reaches its minimum velocity (gray shaded area, Fig. 1C top-left). Top-left plot in Fig. 1C shows the horizontal velocity profiles of the target (solid trace) and reference (dashed trace) disks. On the right, critical motion vectors corresponding to the time at which the target disk's velocity reaches its minimum are shown, with the target disk moving in the same direction as the reference disk but at a lower speed. At the bottom of panel C, the horizontal component of target's instantaneous velocity vector is shown according to different reference frame in each of the five experimental conditions. The spatiotopic motion is identical across all conditions. In the Baseline condition, since only the target is presented, there is no relative motion. However, in all other conditions, the relative motion of the target with respect to the reference disk is in the opposite direction of spatiotopic motion (in this particular example). The retinotopic motion is equal to the spatiotopic motion when observers fixate at the center (i.e. in the Baseline and Fixation conditions). It becomes equal to the relative motion in the SP Same condition whereas the difference between the relative and spatiotopic motions determines its value in the SP Opposite condition. Finally, since observers track a vertically moving object in the SP Orthogonal condition, the *horizontal component* of the retinotopic motion is identical to the horizontal component of retinotopic motion in Baseline and Fixation conditions.

For the particular example described so far, if observers' perception is based solely on the spatiotopic reference frame, they would perceive forward movement for the target disk regardless of condition and report that the target did *not* reverse direction. On the other hand, if their perception is based only on retinal motion, they would report that the target reversed direction in the SP Same and SP Opposite conditions but not in other conditions. As this example demonstrates, observers' tendency to report motion direction reversal depends on how they combine these motion vectors and the aim of this study was to determine this strategy by quantifying the weights applied to each motion vector. Table I summarizes the predicted PSSs in all conditions for each reference frame assuming no influence (i.e. $w = 0$) from the other two.

1.1. Predictions based on a single frame of reference

There are three reference frames available to the observers in all experimental conditions: Retinotopic, spatiotopic, and motion-based frames of reference. Minimum velocities at which subjects perceive the target to be moving backwards (the PSSs) are obtained in all conditions as a function of target-reference center-to-center distance. If perception were solely based on the spatiotopic reference frames, subjects percepts would be independent of experimental conditions (see Fig. 1C) and the PSSs would be zero in all conditions and for all target-reference distances. On the other hand, had observers perceived the direction of motion solely based on retinal motion, backward motion would have been reported whenever the target disk moves backwards on the retina. Therefore, in the Baseline, Fixation, and SP Orthogonal conditions, the PSS would be zero. However, in the SP Same condition in which the reference disk is stationary on the retina (let us assume for the moment that the smooth pursuit gain is 1.0), slightest reduction in speed of the target disk causes backward retinal motion, and hence, the PSS would be equal to the average speed of the target (i.e. 9 deg/s). Moreover, in the SP Opposite condition, according to the retinotopic reference frame, backward motion can only be perceived when the minimum speed of the target disks goes below −9 deg/s. If, as Johansson and many researchers have suggested, the relative motion between the target and reference disks drives perceptual judgments, the PSSs obtained in all conditions (except the Baseline condition in which there is no reference disk) should be 9 deg/s because as soon as the target disk's speed falls below 9 deg/s, it will create a relative motion with respect to the reference disk in the opposite direction of its motion. Predictions of each reference frame in all experimental conditions are summarized in Table I.

Table I
Predictions of PSSs (in deg/s) from spatiotopic (P_s), retinotopic (P_r), and motion-based (P_{mb}) frames of references in different experimental conditions.

	P_s	P_r	P_{mb}
Baseline	0	0	0
Fixation	0	0	9
SP Same	0	9	9
SP Opposite	0	-9	9
SP Orthogonal	0	0	9

2. Methods

2.1. Participants

Three naive observers and one of the authors (MNA) participated in the study. The age of the participants ranged from 26 to 29 years and all participants had normal or corrected-to-normal vision. Experiments followed a protocol approved by the University of Houston Committee for the Protection of Human Subjects and research was carried out in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki). Each observer gave written consent before the experiments.

2.2. Apparatus

Visual stimuli were created via a visual stimulus generator card (VSG2/5, Cambridge Research Systems) and displayed at a resolution of 800×600 with a refresh rate of 100 Hz on a Sony GDM-FW900 CRT monitor. Gaze position monitoring for both eyes was performed by means of an Eyelink-II eye-tracker at 250 Hz sampling rate. The distance between the observer's eyes and the display was 1 m and the dimensions of the display at this distance were 22.7×17.0 deg^2. A head/chin rest was used to help stabilize fixation and to avoid nonlinearities in eye movement recording due to head movements. Observers reported their responses via a joystick.

2.3. Stimuli

Spatial configurations and temporal characteristics of all conditions are given in Fig. 1. White (56 cd/m^2) horizontally moving disks against a black (<0.5 cd/m^2) background were utilized. Experiments were conducted in a normally illuminated room. In some studies, a dark illumination is used when the rationale of the study necessitates a complete elimination of spatiotopic references. However, as discussed in the Introduction, in our study our goal is not to eliminate spatiotopic reference frames. In fact, we are studying the prevailing reference frame in the presence of *all three reference frames,* a situation closer to everyday viewing since under ecological viewing conditions, a static background filled with trees, buildings, or other contextual elements is often present while viewing objects' movements. In order to illustrate that the edges of the display were visible, stimulus configurations shown in Fig. 1 were drawn with gray boundaries. Note also that since observers make motion judgments in the *horizontal* direction, the upper and lower edges of the screen do not provide a reference. The left and right edges do; however, as shown in Fig. 1B, velocity modulations occur only in the central part of the screen where the stimulus is relatively far from the left and right edges of the screen. A fixation point was provided at the beginning of each trial and it was turned off during the motion of the disks to avoid confounding an additional relative motion factor.

We investigated the contributions of three reference frames in perceived motion-direction. Assuming that reference frames combine additively, perceived motion can be modeled as a weighted sum of motion vectors in each reference frame plus a constant term, which represents response bias (Eq. (1)). Since we have four unknowns (three weights and a constant term), we designed four conditions (i.e. four equations) to solve this problem. In order to have a fully determined system of linear equations, the motion of the target disk was represented by a distinct set of vectors according to these reference frames in each condition. In the Baseline condition, a small target disk (0.5 deg) moved horizontally at various vertical eccentricities (2, 5, 8, and 11 deg) with the velocity profile shown in Fig. 1B (for rightward motion). The direction of motion was randomized across trials. Motion duration was 2520 ms. The velocity of both disks was constant during the first and last 630 ms. From 630 ms to 1890 ms, the velocity of the target disk was modulated by a sine wave. Note that as long as the amplitude of the sine wave is smaller than the magnitude of the constant speed component (9 deg/s), the target never moves backward physically (by physically moving backwards, we mean moving backward according to a spatiotopic reference frame, for example, a reference frame centered on the computer monitor). However, if the amplitude of the sine wave exceeds the magnitude of the constant component, the minimum combined velocity of the target will fall below zero and the target will physically move backward (the green arrow in Fig. 1B). In the Fixation condition, the target disk was accompanied by a larger (2.0 deg) disk (referred to as the "reference disk" hereafter) with a constant velocity profile (the dotted horizontal line in Fig. 1B). The presence of a constant-speed reference disk along with the target in the display induces the perception of backward movement in the target, even when the target never moves backward physically (see video demo). The target and reference disks were always located at equal vertical eccentricities from the center of the display but in opposite parts of the display. Which one of the disks is presented in the upper half of the screen was also randomized across trials. In the Baseline and Fixation conditions, observers' eyes remained fixated at the remembered location of the fixation spot (i.e. the center of the screen) throughout the motion of the disks. If the left eye moved outside a 2×2 deg virtual window centered at the center during a trial in these conditions, the trial was discarded and repeated immediately. In the smooth pursuit (SP) Same condition, the reference and target disks had the same velocity profiles shown in Fig. 1B, but observers were asked to pursue the reference disk. In the SP Opposite condition, an additional disk (0.5 deg) with red color, which served as the pursuit target, was presented (Fig. 1A, SP Opposite). In this condition, the pursuit target was always shown at 0 deg vertical eccentricity and moved in the opposite direction (-9 deg/s) of the target and reference disks (the red horizontal line in Fig. 1B). The task of the observers was to report whether the target disk moved back, i.e. the sign of its instantaneous velocity vector has ever changed, at any instant of its motion on the display.

The four conditions described so far were intended to quantify the contribution of three frames of reference: retinotopic, spatiotopic, and motion-based reference frames. After obtaining the weights of each reference frame as a function of distance, in order to test whether they can be generalized to other potential situations as well, we devised another experimental condition as a verification step. In the SP Orthogonal condition (Fig. 1A), the target and reference disks were presented in the same way as in the previous conditions. However, in contrast to the SP Opposite condition, the pursuit target (the red disk) moved vertically. The direction of motion (upward vs. downward) was randomized across trials. The speed of the pursuit target was again 9 deg/s but now in the vertical dimension. The timing of the disk was arranged such that all disks crossed the symmetry axis of the display orthogonal to their motion direction at the same time.

2.4. Procedures

At the beginning of a trial, a fixation spot (0.2 deg) was shown at the center of the screen. Observers were required to press a button on the joystick after establishing proper fixation to carry out drift correction for better accuracy. After drift correction (or 1000 ± 500 ms after the fixation spot was shown in the Baseline and Fixation conditions), the trial started. As soon as all disks completed their motion and disappeared, subjects pressed a button to indicate whether they perceived the target disk to be moving backward or not, in an adaptive staircase design. The amplitude of the sine wave in the target disk's velocity profile was varied by the staircase algorithm across trials. Various realizations of sine modulation are illustrated in Fig. 1B by gray curves. The dependent variable in this study was the "minimum velocity" of the target disk, which corresponds to the dip of the sine wave, or the global minimum in its velocity profile. For instance, a minimum speed of 9 deg/s for the target disk corresponds to 0 deg/s amplitude for the sine wave, i.e. no modulation, whereas a minimum speed of 0 deg/s corresponds to 9 deg/s amplitude for the sine wave. Different target-reference center-to-center distances (vertical separation) were blocked, and each block had four independent staircases interleaved. Each staircase had an initial minimum speed randomly chosen between -10 and 9 deg/s (corresponding to 0–19 deg/s amplitude for the sine wave) and was terminated after ten reversals in subjects' responses (a reversal is a response change from Yes (it moved back), to No (it did not move at all or it moved forward) or vice versa in two consecutive trials within the same staircase). The average of last eight reversals within a staircase was calculated and taken as the point of subjective motion direction reversal or as the point of subjective stationarity (PSS). In other words, staircases converge to a sine wave amplitude which corresponds to the minimum velocity at which the target disk is no longer perceived to be going backwards. As seen from Fig. 1B, if perception were veridical, (i.e. spatiotopic motion on the display were perceived), the PSS would be zero. The minimum step size in the staircase was 0.2 deg/s. A staircase was completed in 15–40 trials depending on the subject and experimental conditions, thus a block of trials could be finished in 60 or as many as 150 trials. The order of blocks (different distances within a given condition) was randomized across subjects. The order of conditions was the same for all subjects (Baseline, Fixation, SP Same, SP Opposite, and SP Orthogonal) and each condition was run on separate days. In order to familiarize the subjects with the stimulus conditions and experimental setup, each subject ran 1 or 2 blocks of trials before collecting data for each condition. One of the practice blocks was always the Baseline trials. This was done to make sure that allocation of attention was similar among conditions. We told observers to spread their attention to the whole display during the experiments to have equivalent allocation of attentional resources among different conditions. However, we cannot completely rule out the use of different strategies in different conditions. In pilot experiments on two observers, we confirmed that the order of conditions and whether staircases for different conditions are interleaved within a block of trials or not, did not affect the results.

3. Results

Fig. 2 shows the PSSs as a function of target-reference center-to-center distance in the Fixation, SP Same, and SP Opposite conditions. Since there was no reference disk in the Baseline, the PSS values in this condition are plotted as a function of vertical eccentricity of the target disk. A two-way repeated measures ANOVA with experimental conditions and distance as the main factors showed significant effect of conditions ($F(3,9) = 11.471$, $p = 0.002$, $\eta_p^2 = 0.793$) and distance ($F(3,9) = 9.656$, $p = 0.004$, $\eta_p^2 = 0.763$). A significant effect for conditions indicates that perception is not veridical and may depend on factors such as eye movements and relative motion. The significant effect of distance confirms previous accounts and shows that the effective reference frame depends on distance. The interaction between condition and distance factors was also significant ($F(9,27) = 4.250$, $p = 0.002$, $\eta_p^2 = 0.586$) implying that distance has different effects in different conditions. Therefore, we carried out one-way repeated measures ANOVAs on each condition to assess the effect of distance in each case. We found a significant effect of distance in the Fixation condition ($F(3,9) = 30.571$, $p < 0.001$, $\eta_p^2 = 0.911$) whereas in the Baseline, SP Same, and SP Opposite conditions, the effect did not reach significance ($F(3,9) = 0.194$, $p = 0.898$, $\eta_p^2 = 0.061$; $F(3,9) = 3.136$, $p = 0.080$, $\eta_p^2 = 0.511$; $F(3,9) = 2.985$, $p = 0.089$, $\eta_p^2 = 0.499$, respectively). We also fitted linear regression lines to the results in each condition to get quantitative measures of how much distance affects the PSSs. In the Baseline condition, the percepts are veridical (i.e. the PSSs are around zero), and vertical eccentricity does not have any effect (slope = 0.010; intercept = −0.580). In the Fixation condition, for each degree of separation between the target and reference disks, the effect size drops by 0.4 deg/s (slope = −0.400; intercept = 6.170). In the SP Same condition, interestingly, the effect reaches 90% (intercept = 8.520) of the physical speed and becomes immune to changes in target-disk separation (slope = −0.020) within the range of distances used in this study. This may be due to the fact that retinotopic and motion-based frames of references reinforce the same percept (both predict an effect size of 9 deg/s). In the SP Opposite condition, the effect of distance is reduced. Nevertheless, the distance is marginally significant (slope = −0.293, $p = 0.059$; intercept = 6.7 34).

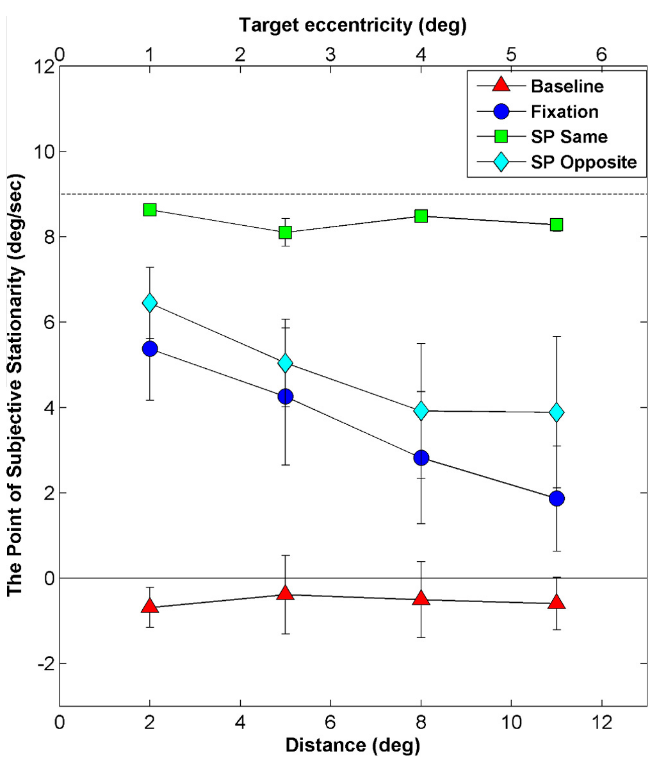

Fig. 2. The points of subjective stationarity in all conditions are plotted against center-to-center distance between the target and reference disks. Horizontal dotted line represents the average velocity of the target disk and the constant velocity of the reference disk. The secondary x axis represents the vertical eccentricity of the target disk. Error bars indicate ± SEM ($n = 4$).

We also carried out regression fits for each subject, individually. Fig. 3 shows the slopes and intercepts from all subjects in all experimental conditions. Different marker shapes (and colors) represent various experimental conditions. Each point represents data from a single subject and error bars are the standard errors of the parameter estimations. The vertical dashed line represents the average physical speed, i.e. the ceiling for the perceptual effect. Note that all the points in Fig. 5 fall on the left of this physical limit, i.e. intercepts smaller than 9 deg/s. This finding itself lends support to the hypothesis that each reference frame contributes to the percept with varying weights. Furthermore, most of the points fall below the horizontal axis, which is indicative of distance-dependent contribution of at least some of the reference frames. Data from different subjects for each condition form clusters with few exceptions. Nevertheless, subject-to-subject differences do not prevent generalization of the slope and intercept values to the population of observers. Therefore, the following analyses are done on average data.

3.1. Quantifying the weight of each frame of reference

We fitted Eq. (1) to data in Fig. 2 and estimated the weights of each reference frame at various target-reference distances. So far, we have assumed perfect fixation stability and ideal smooth-pursuit gain (i.e., equal to 1) in outlining the predictions of each reference frame. However, in practice, subjects often make micro-saccades and drifts during fixation. Stability worsens especially when there is no fixation target during stimulus presentation, as it was the case in our experiments. In addition, smooth pursuit gains are generally lower than 1, which may lead to overestimation of predictions of the retinotopic reference-frame. For example,

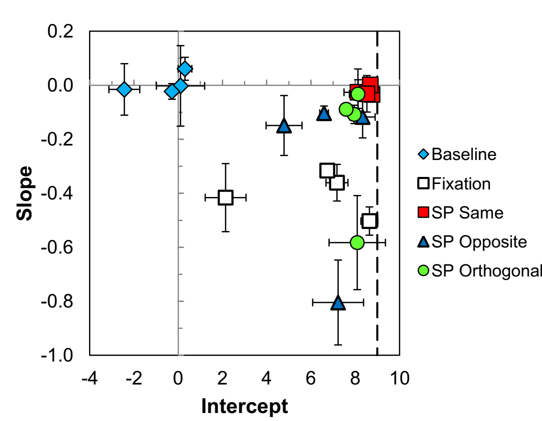

Fig. 3. Slope and intercept values obtained by linear regression of the PSSs obtained in different conditions. Each marker represents a single observer. Different marker types (or colors) represent different conditions. Error bars show ± SEM over four randomly interleaved staircases. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

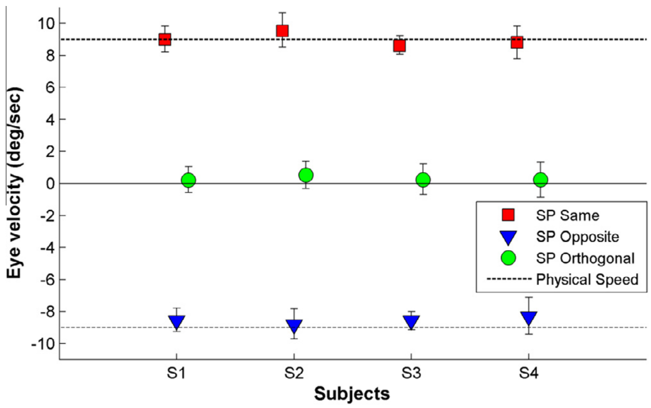

Fig. 4. Median eye velocities in three different experimental conditions in which observers were asked to do smooth pursuit. Error bars represent one interquartile range. Dotted horizontal lines represent physical velocities of the reference disk in the SP Same and SP Opposite conditions respectively.

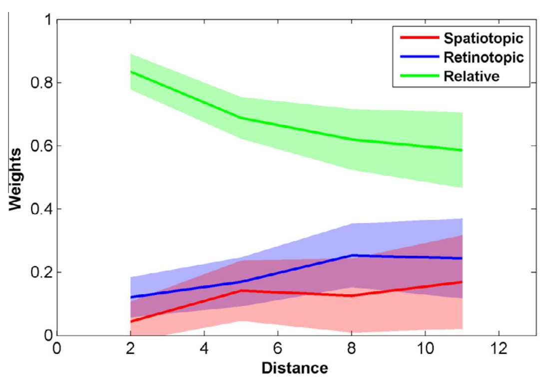

Fig. 5. Weights of retinotopic, spatiotopic, and motion-based frames of reference in determining the perceptual judgments of motion direction. The *X* axis is distance, and the *Y* axis represents the weights. Shaded regions around the curves represent standard errors estimated by bootstrapping the residuals 300 times.

assume that a subject has a smooth pursuit gain of 0.7. This means that the retinal speed of the reference disk will not be 0 but $(1-0.7) \times 9 = 2.7$ deg/s. This would decrease the PSS. In order to take these into account, we measured eye movements during the experiments. Fig. 4 shows the median horizontal eye velocities from each subject in smooth pursuit experiments. Error bars represent one interquartile range from medians. Eye movement speeds in the SP Same and Opposite conditions were not significantly different from the physical speed of the pursuit target (Note that the direction of pursuit was in the opposite direction in SP Opposite condition). Horizontal eye velocities in the SP Orthogonal condition were not significantly different than zero for all subjects. We fine-tuned the predictions of each reference frame given in Table I by using average eye velocities recorded during smooth pursuit and fixation. Fig. 5 shows the weights of each reference frame as a function of distance. The *X* and *Y* axes represent the distance and the weights, respectively. Different marker types (and colors) indicate different reference frames (retinotopic, spatiotopic, and motion-based). At all distances, the motion-based reference frame dominates such that the relative speed of the target disk with respect to the reference disk is perceived as backward motion even though the physical speed of the target disk was always in the same direction. There is a clear drop in the weight of the motion-based reference frame with increasing distance. Moreover, retinotopic and spatiotopic reference frames have similar weights. Their weights stay relatively constant with increasing distance.

3.2. Putting the weights to the test

We have quantified the weight of each reference frame as a function of distance, i.e. $w(d)$. In order to test whether the perceptual system uses the same weight functions in other situations as well, we designed another experimental condition, namely the SP Orthogonal condition, in which the target and the reference disks moved as in previous conditions while the pursuit target (the red disk) moved vertically. Since horizontal eye movements are assumed to be negligible during vertical pursuit, this condition leads to predictions very similar, if not identical, to those in the Fixation condition. Assuming perfect vertical pursuit without any horizontal component in the eye movements, the prediction coefficients in Eq. (1) for the SP Orthogonal condition are $P_s = 0$, $P_r = 0$ and $P_{mb} = 9$ deg/s; in other words, identical to those in the Fixation condition (see Table I). Imperfect pursuit performance will only affect P_r, which would then be equal to the average horizontal speed during vertical pursuit. The only difference between the two conditions is the existence of an ongoing vertical eye movement in the SP Orthogonal condition. Data from this experiment are shown in Fig. 6. A two-way repeated measures ANOVA with main factors of distance and conditions (the Fixation vs. the SP Orthogonal) showed no significant difference between the two conditions ($F(1,3) = 2.343$, $p = 0.223$, $\eta_p^2 = 0.439$). This allows us to test the weights we estimated previously in the SP Orthogonal condition. Distance had a significant effect ($F(3,9) = 10.949$, $p = 0.002$, $\eta_p^2 = 0.785$). The interaction between conditions and distance was not significant ($F(3,9) = 2.015$, $p = 0.182$, $\eta_p^2 = 0.402$). Linear regression resulted in an intercept of 7.939 and a slope of -0.203.

We used the predictions and the weight functions estimated from the other four conditions to predict the results in the SP Orthogonal condition (Fig. 3 also shows individual slopes and intercepts in the SP Orthogonal condition). Fig. 6 shows experimental data (markers) and the model predictions (solid line). The model fit follows a similar pattern with the data. In fact, bootstrap-

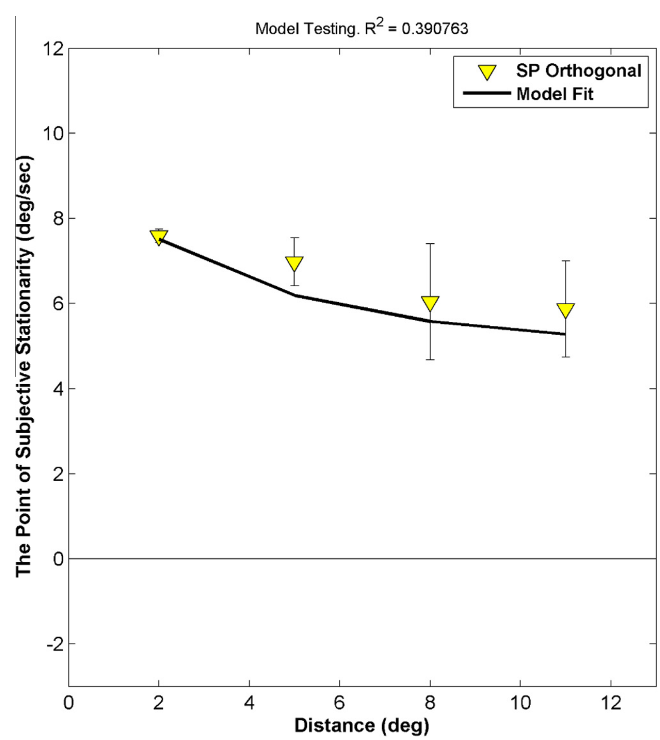

Fig. 6. Model fit and experimental data in the SP Orthogonal condition. Error bars represent SEM ($n = 4$).

ping for paired samples *t*-test resulted in no significant difference between the model fit and the experimental data ($t(3) = 0.555$, $p = 0.618$, $d = 0.377$). However, there is a slight underestimation of the effect size, which is reflected by the coefficient of determination ($R^2 = 0.391$).

4. Discussion

The aim of this study was to investigate the strategies used by the perceptual system to cope with the dynamic changes in the retinal stimulation. More specifically, we sought to understand the way various reference frames are utilized for motion computations. We used a variant of the induced motion paradigm, and in four conditions, we measured how direction of motion of a target disk is perceived when there are several reference frames that could be utilized in motion calculations. Specifically, we quantified the contributions of retinotopic, spatiotopic, and motion-based reference frames in a simple model. We found that the relative motion of the target disk with respect to the reference disk (i.e. motion-based reference frame) outweighs the other two, and its effectiveness is dependent on the distance between the target and the reference disks. This finding suggests that relative motion constitutes a significant part of the percepts within the range of distances tested here (up to 11 deg). Next, we ran another condition to see whether these partial contributions are specific to the first four conditions, or they generalize to other possible situations as well. The predictions of each reference frame was the same in the Fixation and the SP Orthogonal conditions, allowing us to directly test the performance of our model. Results in these two experiments were not statistically different from each other, implying that the use of the proposed model is warranted. The proposed model successfully predicted the experimental results in the SP Orthogonal condition. There was, though, a small underestimation of the effect size, which suggests that there might be additional factors that influence perceptual judgments of motion direction. Nevertheless, the weighted combination of retinotopic, spatiotopic, and motion-based reference frames performed reasonably well in explaining the percepts.

We do not, however, overgeneralize these findings because there are still undetermined factors, which became evident with small underestimation of the effect size in the SP Orthogonal condition. One of the potential factors may be the mere existence of an ongoing eye movement. In other words, the presence of ongoing eye movements might have influenced motion detection mechanisms in a complex manner. During self-motion or eye movements, the perceptual system may rely more on relative motions. Ecologically speaking, while an organism is on the move in its habitat or moves its eyes, retinal images undergo global changes and relative motion of objects might be more relevant for survival. Therefore, it is a reasonable strategy to put more weight on relative changes during eye movements or self-motion of an observer. However, as statistical tests confirmed, we did not find any difference between the Fixation and SP Orthogonal conditions, conditions in which the only difference was the mere existence of eye movements in the latter. Contrasting these two conditions show that the sole existence of an ongoing motor action is not the missing factor. Allocation of spatial attention during fixation vs. smooth pursuit may also be different. Gogel and Sharkey (1989) measured the perceived motion trajectory of a vertically moving object in the presence of one or two objects (i.e. inducers) moving horizontally. Observers were tracking the vertically moving object. In the case of a single inducer, they found a substantially larger tilt in the trajectory of the tracked object when the inducer spot was attended than ignored. Similarly, when there were two inducer spots moving horizontally in opposite directions, attending to one of the inducers clearly increased effectiveness of the attended object. In our experiments, we asked observers to spread their attention to the whole display; however, we did not have any control over allocation of spatial attention on a trial-by-trial basis. Therefore, observers might have used various strategies to allocate their attention. This point needs further experimentation.

4.1. The effective reference frame in perception

In general, the choice of a reference frame cannot be done at will, and is neither a result of an all-or-none process nor an outcome of a winner-take-all competition among neural representations. Each reference frame exerts its effect with varying weights depending on its relevance and the spatio-temporal characteristics of the retinal stimuli. There are several studies supporting partial contributions of different exogenous (space-based, object-based, motion-based, etc.) reference frames depending on the spatial structure and geometrical organization (Farrell-Whelan & Brooks, 2013; Magnussen, Orbach, & Loffler, 2013; Shum & Wolford, 1983; Tadin, Lappin, Blake, & Grossman, 2002), depth (Gogel, 1974; Gogel & Koslow, 1972), belongingness (DiVita & Rock, 1997), speed (Hisakata et al., 2013; Léveillé & Yazdanbakhsh, 2010; Mori, 1984), lighting conditions (Shum & Wolford, 1983), eccentricity (Thurman & Lu, 2013), and even interactions among different modalities (Avillac, Denève, Olivier, Pouget, & Duhamel, 2005). Furthermore, studies on perception during voluntary movement of eyes, head, or body indicate varying contributions of endogenous (retinocentric, headcentric, etc.) and exogenous reference frames (Agaoglu, Ogmen, & Herzog, 2012; Becklen et al., 1984; Brenner & van den Berg, 1994; Durgin, Gigone, & Scott, 2005; Hisakata et al., 2013; Johansson, 1976; Souman et al., 2006a; Turano & Heidenreich, 1999).

Durgin et al. (2005) measured perceived speeds of visual flow under the influence of these factors, i.e. while walking on a treadmill, during physical translation (without biomechanical self-motion), and during normal walking. They found that each factor (walking without translation and translation without walking) reduces the perceived speed, and they approximately add up, i.e. the reduction is greatest during normal walking (Durgin et al., 2005). Brenner and van den Berg (1994) asked subjects to pursue a target moving against a textured background. In the midst of its trajectory, the target could increase or decrease its velocity but the background texture could also move with the target. Subjects were asked to indicate whether the target moved faster, at the same speed, or more slowly during the final interval than it had in the initial interval. They found that, as long as the target-background relative motion is kept constant, i.e. the retinal slip of the background is in the opposite direction of the eye movement, perceived target velocity does not change even if the physical speed of the target is increased or decreased. In addition, when the target speed specified by retinal signals is slower than what extra-retinal signals indicate, or is in the opposite direction, extra-retinal signals dictate perceived speed judgments (Brenner & van den Berg, 1994). Hisakata et al. (2013) investigated motion-induced position shifts during smooth pursuit eye movements. In different conditions, they varied the motion of the carrier of a moving Gabor patch to pin down the critical reference frame for the illusion to occur. They found that the illusion occurs according to the envelope-relative velocity of the carrier (Hisakata et al., 2013). In other words, a motion-based reference frame drives the illusion, which again emphasizes the dominance of relative motion over retinotopic and spatiotopic motions. Anstis and Casco also demonstrated in the "flying bluebottle illusion" (see also Furman & Gur 2005; Kano & Hayashi, 1981) that relative motion drives perception, which is reflected in shape and size judgments in their experiments (Anstis & Casco, 2006). However, the effect sizes they obtained were much larger than what would be expected by just relative motion between objects and background. This indicates that there are still other factors that might shape percepts, which is also in line with what we have found in our study. Of course, studies demonstrating the dominance of relative motion are not limited to those cited here. All these studies imply that the visual system, although it is mostly organized retinotopically in its early areas (Sereno et al., 1995; Tootell, Silverman, Switkes, & De Valois, 1982), chooses alternative representations over retinotopic ones, as necessitated from an ecological point of view. How does, then, the visual system choose the effective reference frame for perception? Is this process determined by some parameters of the visual stimuli or is it a result of selective allocation of attention on specific features?

4.2. Non-retinotopic processes

Dynamic changes in the environment and the movement of the eyes, head, and body necessitate representations of objects and events (changes over time) that are invariant to such changes. Indeed, there have been many psychophysical (Agaoglu et al., 2012; Boi, Ogmen, Krummenacher, Otto, & Herzog, 2009; Boi, Vergeer, Ogmen, & Herzog, 2011; Kawabe, 2008; Léveillé, Myers, & Yazdanbakhsh, 2014; Nishida, Watanabe, Kuriki, & Tokimoto, 2007; Ogmen, Otto, & Herzog, 2006; Otto, Ogmen, & Herzog, 2010; Shimozaki, Eckstein, & Thomas, 1999), functional imaging (Galati et al., 2000; Maus, Fischer, & Whitney, 2013; Neggers, Van der Lubbe, Ramsey, & Postma, 2006; Yin, Shimojo, Moore, & Engel, 2002; Zaehle et al., 2007), and neurophysiological studies (Bremner & Andersen, 2014; Moorman & Olson, 2007; Olson, 2003) on the existence of robust non-retinotopic processes and representations. Since early visual areas are organized retinotopically (Sereno et al., 1995; Tootell et al., 1982), these studies indirectly suggest that representations based on multiple reference frames are constructed by different levels of processing in the brain. In fact, it has been shown that a spectrum of neural representations of endogenous (e.g. retinocentric motion) and exogenous (e.g. relative motion, optic flow, etc.) motions coexist in the brain (Arnoldussen, Goossens, & van den Berg, 2011; Avillac et al., 2005; Colby & Goldberg, 1999; Inaba, Shinomoto, Yamane, Takemura, & Kawano, 2007; Malkinson, McKyton, & Zohary, 2012; Neggers et al., 2006; Takemura, Ashida, Amano, Kitaoka, & Murakami, 2012). Under some conditions, the observer may be able to combine different reference frames according to task demands; however, in many cases, it is not possible for the observer to pick a reference frame at will (try for example to perceive a stationary object in motion according to a retinotopic reference frame while you are moving your eyes). A variety of factors such as adjacency, similarity, belongingness, center of gravity, figural organization, and attention, influence how reference frames are selected or combined. Further, when all parameters are kept the same, the same stimuli could be perceived differently by different observers. For instance, the classical experiment on induced motion with two dots (one moving, one stationary) showed that some observers perceive motion on one dot, some attribute motion to the stationary one, and some perceive both dots as moving at intermediate values (Day, 1978; Mack, Fisher, & Fendrich, 1975; Wallach, 1959).

4.3. Implications for perceptual vector decomposition

Distance dependent changes in motion perception, like we have shown in this study, have been attributed to imperfect extraction of common motion vectors when there are multiple objects in a dynamic scene (Hochberg & Fallon, 1976; Shum & Wolford, 1983). These studies have shown a finite spatial distance between moving dots, in which there is a linear drop in extracted common-motion components. On the contrary, Johansson in his *theory of perceptual vector decomposition*, implicitly claimed that extraction of common motion is complete (Johansson, 1950, 1973, 1976). For instance, he explained the rotary motion of the peripheral light in Duncker's wheel stimuli by perceptual subtraction of the common motion vector from its cycloidal motion (Johansson, 1950, 1976). Johansson also demonstrated a unique ability of the human perceptual system with his biological motion displays (Johansson, 1973, 1976). Constructed by only 5–10 point-lights placed at the joints of an otherwise invisible actor, biological motion displays contain highly complex motion patterns with respect to a retinotopic or spatiotopic reference frame. Interestingly, observers can still clearly identify the type of motion even if stimulus duration is very short, e.g. 0.1 s (Johansson, 1973). Bardi, Regolin, and Simion (2011) provided evidence supporting the view that this ability is inborn in humans by demonstrating preference of newborns to look at biological motion displays. Johansson attributed this ability to perceptual vector decomposition. In fact, with his theory of vector analysis, biological motion displays could be described by a hierarchy of moving reference frames, thus simplifying the motions of knees and feet as simple harmonic motion of a pendulum (Johansson, 1973, 1976). Supporting his suggestions, other examples of hierarchical reference frames and potential neural processes that give rise to decomposition of these reference frames have been investigated (Bertamini & Proffitt, 2000; Grossberg, Léveillé, & Versace, 2011; Sokolov & Pavlova, 2006). Following the Gibsonian approach to the problem of perceptual stability (Gibson, 1979), Johansson also proposed that frequent changes in the proximal stimuli (due to motion of the eyes, head, or body, and motion of objects in the environment) do not pose any problems because the visual system is tuned to "abstracting" information from change in retinal stimuli over time (Jansson, Bergstrom, & Epstein, 1994; Johansson, von Hofsten, & Jansson, 1980). This abstraction involves extracting common motion in three dimensions and using it as a reference frame to represent other changes in the retinal flow. This way, Johansson argued that motion resulting from self-movement can be distinguished from the motion of objects in the environment.

Johansson's theory of perceptual vector analysis (Johansson, 1973) had three principles: (i) elements under motion are always perceptually related, (ii) simultaneous motion of elements form rigid perceptual groups, (iii) the decomposition of motion vectors into equal and simultaneous motion vectors leads to the perception of "common motion", and the residual motion vectors will be perceived as "relative motion". Although this seems logical, biological implementation of vector decomposition is not as simple as it appears. First, the extraction of common motion vectors may not always be the same for a given combination of point lights. For instance, when presented with the Duncker's wheel stimuli, while some observers reported a rotating wheel, others reported that the motion of two point lights resembled more to a tumbling stick (Cutting & Proffitt, 1982; Proffitt, Cutting, & Stier, 1979). Different speeds along the same motion trajectories have been shown to result in both subject-to-subject and trial-to-trial differences in perceived common motion in two point light displays (Mori, 1984). More importantly, in mathematical terms, vector decomposition is an ill-posed problem: Infinitely many pairs of common and relative motions can produce exactly the same absolute motion, i.e. motion with respect to a stationary exogenous reference frame. Therefore, a fundamental question in vector decomposition has been to determine which of the infinitely many solutions is adopted by the visual system.

Several heuristics have been proposed as to how vector decomposition takes place in the brain. Here, we describe only one of them in detail. In Wade and Swanston's (1987) formulation, "...

finding the common motion is equivalent to a comparison of pairs of relational [i.e. relative] *motions, and selecting those which have the same value*". And "*it seems reasonable to adopt those points that have the same relational motion values as a patterncentric frame of reference for allocating other relational motions*" (Wade & Swanston, 1987, p. 564). According to the proposed method, the perceptual system has to do pairwise comparisons for all the points (objects or features). If this process were to be carried out in an iterative way, it would run into combinatorial explosion in a natural environment. A parallel implementation of this process would require a complex neural architecture that needs to be spelled out.

4.4. An alternative to vector decomposition: The Reference-Frame Metric Field theory

Recently, we proposed the Reference-Frame Metric Field (RFMF) theory (Ogmen, Herzog, & Noory, 2013) in order to study how reference frames are established and how dynamic computations of form is carried out. Here we describe only the part of RFMF that is relevant to this study. Fig. 7 provides a schematic description of the RFMF theory. At the bottom, the retinotopic space is illustrated. Here, several simple stimuli (dots) are in motion. These motion vectors are grouped locally to generate local motion vectors. These local motion vectors serve as *local reference frames*. According to our theory, *a field is created around each local reference frame* (like an electromagnetic field) and fields of different reference frames *interact* to establish a global equilibrium in the retinotopic space. Therefore, the visual system does not need to solve explicitly the ill-posed common motion extraction problem. Instead, motion-based reference frames result from a process of *reference field* interactions. Interactions of various fields give rise to perceptual organizations such that different frames of reference may dominate different regions in the perceptual space. The degree to which extraction process takes place may be incomplete (i.e. the perceived and physical common motion vectors may not be equal) and can vary with spatio-temporal properties of the stimuli (DiVita & Rock, 1997; Gogel & Koslow, 1972; Johansson et al., 1980; Mori, 1979, 1984; Poljac, Verfaillie, & Wagemans, 2011; Shum & Wolford, 1983). This finding rejects perfect vector decomposition and rather calls for a finite region within which the extraction of common motion component can be carried out.

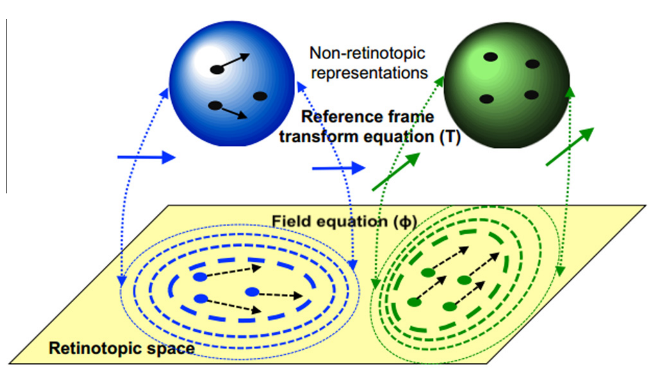

Fig. 7. Schematic illustration of the RFMF theory. Motions of several blue and green colored dots are shown in the retinotopic space (the yellow plane). Dots are grouped into two groups based on their motion vectors. A reference motion vector is extracted and serves as the reference by which dots are mapped into non-retinotopic representations (blue and green spheres at the top). Reference motion vectors' effect spread over space and time much like an electromagnetic field (dotted ellipses whose thicknesses symbolise field strength). Interactions between different reference fields, if there are many, determine the resultant non-retinotopic representations. For more details, see (Ogmen, 2007; Ogmen & Herzog, 2010). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Yet within this finite region many moving stimuli can exist, each providing a potential reference frame. This region might depend on stimulus scale (Maruya, Holcombe, & Nishida, 2013) and other parameters which need further investigation. In terms of neural implementation, field interactions can be realized by distance-dependent isotropic (in case of no direction bias) or anisotropic (in case of direction bias) connections among units carrying out the computations.

5. Conclusion

Taken together, our results suggest that the effective reference frame for motion perception involves a combination of motion-based, retinotopic and spatiotopic reference frames. Relative motions defined by motion-based reference frames dominate retinotopic and spatiotopic motions. Effectiveness of the motion-based reference frames drops substantially as the distance between objects of interest increases, indicating a finite region within which each motion-based frame of reference operates. Such dramatic changes are not found for retinotopic and spatiotopic reference frames. Contributions of retinotopic and spatiotopic frames of reference are minimal when there is an ongoing smooth pursuit eye movement. From the perspective of the RFMF theory, motion-based reference frames emerge from field-like interactions of local motion vectors, thereby providing an alternative account to the vector decomposition approach. Distance-dependent changes in perceived motion result from interactions between multiple motion-based reference frames. These interactions determine an effective reference frame whereby information from retinotopic representations can be mapped into non-retinotopic ones. In our previous studies, we have suggested that such a mapping allows dynamic computation of form while avoiding motion blur, moving ghosts, and occlusion problems (Ogmen, 2007; Ogmen & Herzog, 2010).

Acknowledgment

Michael Herzog is supported by the Swiss National Science Foundation (SNF) project "Basics of visual processing: from retinotopic encoding to non-retinotopic representations."

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.visres.2014.12.009.

References

Agaoglu, M. N., Ogmen, H., & Herzog, M. H. (2012). Non-retinotopic feature processing in the absence of retinotopic spatial layout and the construction of perceptual space from motion. *Vision Research, 71*, 10–17.
Anstis, S. M., & Casco, C. (2006). Induced movement: The flying bluebottle illusion. *Journal of Vision, 6*(10), 1087–1092.
Arnoldussen, D. M., Goossens, J., & van den Berg, A. V. (2011). Adjacent visual representations of self-motion in different reference frames. *Proceedings of the National Academy of Sciences of the United States of America, 108*, 11668–11673.
Aubert, H. (1886). Die Bewegungsempfindung. *Pflügers Archiv, 39*, 347–370.
Avillac, M., Denève, S., Olivier, E., Pouget, A., & Duhamel, J.-R. (2005). Reference frames for representing visual and tactile locations in parietal cortex. *Nature Neuroscience, 8*(7), 941–949.
Baker, C. J., & Braddick, O. (1982). Does segregation of differently moving areas depend on relative or absolute displacement? *Vision Research, 22*(7), 851–856.
Bardi, L., Regolin, L., & Simion, F. (2011). Biological motion preference in humans at birth: Role of dynamic and configural properties. *Developmental Science, 14*(2), 353–359.
Becklen, R., Wallach, H., & Nitzberg, D. (1984). A limitation of position constancy. *Journal of Experimental Psychology. Human Perception and Performance, 10*(5), 713–723.

Bertamini, M., & Proffitt, D. R. (2000). Hierarchical motion organization in random dot configurations. *Journal of Experimental Psychology. Human Perception and Performance, 26*(4), 1371–1386.
Boi, M., Ogmen, H., Krummenacher, J., Otto, T. U., & Herzog, M. H. (2009). A (fascinating) litmus test for human retino-vs. non-retinotopic processing. *Journal of Vision, 9*, 1–11.
Boi, M., Vergeer, M., Ogmen, H., & Herzog, M. H. (2011). Nonretinotopic exogenous attention. *Current Biology: CB, 21*(20), 1732–1737.
Bremner, L. R., & Andersen, R. A. (2014). Temporal analysis of reference frames in parietal cortex area 5d during reach planning. *Journal of Neuroscience, 34*(15), 5273–5284.
Brenner, E., & van den Berg, A. V. (1994). Judging object velocity during smooth pursuit eye movements. *Experimental Brain Research, 99*(2), 316–324.
Colby, C. L., & Goldberg, M. E. (1999). Space and attention in parietal cortex. *Annual Review of Neuroscience, 22*, 319–349.
Cutting, J. E., & Proffitt, D. R. (1982). The minimum principle and the perception of absolute, common, and relative motions. *Cognitive Psychology, 14*(2), 211–246.
Day, R. (1978). Induced visual movement as nonveridical resolution of displacement ambiguity. *Perception & Psychophysics, 23*(3), 205–209.
DiVita, J. C., & Rock, I. (1997). A belongingness principle of motion perception. *Journal of Experimental Psychology. Human Perception and Performance, 23*(5), 1343–1352.
Dodge, R. (1904). The participation of the eye movements in the visual perception of motion. *Psychological Review, 11*(1), 1–14.
Duffy, C. J., & Wurtz, R. H. (1991a). Sensitivity of MST neurons to optic flow stimuli. I. A continuum of response selectivity to large-field stimuli. *Journal of Neurophysiology, 65*(6), 1329–1345.
Duffy, C. J., & Wurtz, R. H. (1991b). Sensitivity of MST neurons to optic flow stimuli. II. Mechanisms of response selectivity revealed by small-field stimuli. *Journal of Neurophysiology, 65*(6), 1346–1359.
Duncker, K. (1929). Über induzierte Bewegung. *Psychologische Forschung, 12*(1), 180–259. http://dx.doi.org/10.1007/BF02409210.
Durgin, F. H., Gigone, K., & Scott, R. (2005). Perception of visual speed while moving. *Journal of Experimental Psychology. Human Perception and Performance, 31*(2), 339–353.
Farrell-Whelan, M., & Brooks, K. R. (2013). Differential processing: Towards a unified model of direction and speed perception. *Vision Research, 92*, 10–18.
Festinger, L., Sedgwick, H. A., & Holtzman, J. D. (1976). Visual perception during smooth pursuit eye movements. *Vision Research, 16*(12), 1377–1386.
Filehne, W. (1922). Uber das optische Wahrnehmen von Bewegungen. *Zeitschrift für Sinnephysiologie, 53*.
Fleischl, E. V. (1882). Physiologisch-optische Notizen, 2. Mitteilung. *Sitzung Wiener Bereich der Akademie der Wissenschaften, 3*, 7–25.
Freeman, T. C. A. (2001). Transducer models of head-centred motion perception. *Vision Research, 41*(21), 2741–2755.
Freeman, T. C. A., & Banks, M. S. (1998). Perceived head-centric speed is affected by both extra-retinal and retinal errors. *Vision Research, 38*(7), 941–945.
Freeman, T. C. A., Champion, R. A., Sumnall, J. H., & Snowden, R. J. (2009). Do we have direct access to retinal image motion during smooth pursuit eye movements? *Journal of Vision, 9*(33), 1–11.
Freeman, T. C. A., Champion, R. A., & Warren, P. A. (2010). A Bayesian model of perceived head-centered velocity during smooth pursuit eye movement. *Current Biology: CB, 20*(8), 757–762.
Furman, M., & Gur, M. (2003). Self-organizing neural network model of motion processing in the visual cortex during smooth pursuit. *Vision Research, 43*(20), 2155–2171.
Furman, M., & Gur, M. (2005). Alteration of the perceived path of a non-pursued target during smooth pursuit: Analysis by a neural network model. *Vision Research, 45*(13), 1755–1768.
Galati, G., Lobel, E., Vallar, G., Berthoz, A., Pizzamiglio, L., & Le Bihan, D. (2000). The neural basis of egocentric and allocentric coding of space in humans: A functional magnetic resonance study. *Experimental Brain Research, 133*(2), 156–164.
Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston: Houghton Mifflin.
Gogel, W. C. (1974). Relative motion and the adjacency principle. *Quarterly Journal of Experimental Psychology, 26*(3), 425–437.
Gogel, W. C. (1977). The metric of visual space. In W. Epstein (Ed.), *Stability and constancy in visual perception, mechanisms, and processes* (pp. 129–181). New York: John Wiley.
Gogel, W. C., & Koslow, M. (1972). The adjacency principle and induced movement. *Perception & Psychophysics, 11*(4), 309–314.
Gogel, W. C., & Sharkey, T. (1989). Measuring attention using induced motion. *Perception, 18*(3), 303–320.
Grossberg, S., Léveillé, J., & Versace, M. (2011). How do object reference frames and motion vector decomposition emerge in laminar cortical circuits? *Attention, Perception & Psychophysics, 73*(4), 1147–1170.
Hisakata, R., Terao, M., & Murakami, I. (2013). Illusory position shift induced by motion within a moving envelope during smooth-pursuit eye movements. *Journal of Vision, 13*(21), 1–12.
Hochberg, J. E., & Fallon, P. (1976). Perceptual analysis of moving patterns. *Science, 194*(4269), 1081–1083.
Inaba, N., Shinomoto, S., Yamane, S., Takemura, A., & Kawano, K. (2007). MST neurons code for visual motion in space independent of pursuit eye movements. *Journal of Neurophysiology, 97*, 3473–3483.
Jansson, G., Bergstrom, S., & Epstein, W. (1994). *Perceiving events and objects*. Hillsdale, NJ: Lawrence Erlbaum Associates.
Johansson, G. (1950). *Configurations in event perception: An experimental study*. Stockholm: Almqvist & Wiksell.
Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics, 14*(2), 201–211.
Johansson, G. (1976). Spatio-temporal differentiation and integration in visual motion perception. *Psychological Research, 38*(4), 379–393.
Johansson, G., von Hofsten, C., & Jansson, G. (1980). Event perception. *Annual Review of Psychology, 31*, 27–63.
Kano, C., & Hayashi, K. (1981). The apparent path of a stationary and a circularly moving spot during the smooth pursuit of another circularly moving spot. *Acta Psychologica, 48*(1–3), 151–160.
Kawabe, T. (2008). Spatiotemporal feature attribution for the perception of visual size. *Journal of Vision, 8*(8), 7.1–9.
Léveillé, J., Myers, E., & Yazdanbakhsh, A. (2014). Object-centered reference frames in depth as revealed by induced motion. *Journal of Vision, 14*(3), 1–11.
Léveillé, J., & Yazdanbakhsh, A. (2010). Speed, more than depth, determines the strength of induced motion. *Journal of Vision, 10*, 1–9.
Mack, A., Fisher, C., & Fendrich, R. (1975). A reexamination of two-point induced movement. *Perception & Psychophysics, 17*(3), 273–276.
Mack, A., & Herman, E. (1972). A new illusion: The underestimation of distance during pursuit eye movements. *Perception & Psychophysics, 12*(6), 471–473.
Mack, A., & Herman, E. (1973). Position constancy during pursuit eye movement: An investigation of the Filehne illusion. *The Quarterly Journal of Experimental Psychology, 25*(1), 71–84.
Mack, A., & Herman, E. (1978). The loss of position constancy during pursuit eye movements. *Vision Research, 18*(1), 55–62.
Magnussen, C. M., Orbach, H. S., & Loffler, G. (2013). Motion trajectories and object properties influence perceived direction of motion. *Vision Research, 91*, 21–35.
Malkinson, T., McKyton, A., & Zohary, E. (2012). Motion adaptation reveals that the motion vector is represented in multiple coordinate frames. *Journal of Vision, 12*, 1–11.
Maruya, K., Holcombe, A., & Nishida, S. (2013). Rapid encoding of relationships between spatially remote motion signals. *Journal of Vision, 13*, 1–20.
Mateeff, S., & Hohnsbein, J. (1989). The role of the adjacency between background cues and objects in visual localization during ocular pursuit. *Perception, 18*(1), 93–104.
Mateeff, S., Hohnsbein, J., & Ehrenstein, W. H. (1990). Visual localization and estimation of extent of target motion during ocular pursuit: A common mechanism? *Perception, 19*(4), 459–469.
Maus, G. W., Fischer, J., & Whitney, D. (2013). Motion-dependent representation of space in area MT+. *Neuron, 78*(3), 554–562.
Moorman, D. E., & Olson, C. R. (2007). Combination of neuronal signals representing object-centered location and saccade direction in macaque supplementary eye field. *Journal of Neurophysiology, 97*(5), 3554–3566.
Mori, T. (1979). Relative locations among moving spots and visual vector analysis. *Perceptual and Motor Skills, 48*(2), 587–592.
Mori, T. (1984). Change of a frame of reference with velocity in visual motion perception. *Perception & Psychophysics, 35*(6), 515–518.
Morrone, M. C., Tosetti, M., Montanaro, D., Fiorentini, A., Cioni, G., & Burr, D. (2000). A cortical area that responds specifically to optic flow, revealed by fMRI. *Nature Neuroscience, 3*(12), 1322–1328.
Neggers, S. F. W., Van der Lubbe, R. H. J., Ramsey, N. F., & Postma, A. (2006). Interactions between ego- and allocentric neuronal representations of space. *NeuroImage, 31*(1), 320–331.
Nishida, S. (2011). Advancement of motion psychophysics: Review 2001–2010. *Journal of Vision, 11*, 1–53.
Nishida, S., Watanabe, J., Kuriki, I., & Tokimoto, T. (2007). Human visual system integrates color signals along a motion trajectory. *Current Biology: CB, 17*(4), 366–372.
Ogmen, H. (2007). A theory of moving form perception: Synergy between masking, perceptual grouping, and motion computation in retinotopic and non-retinotopic representations. *Advances in Cognitive Psychology, 3*(1–2), 67–84.
Ogmen, H., & Herzog, M. H. (2010). The geometry of visual perception: Retinotopic and non-retinotopic representations in the human visual system. *Proceedings of the IEEE Institute of Electrical and Electronics Engineers, 98*(3), 479–492.
Ogmen, H., Herzog, M. H., & Noory, B. (2013). Reference-frame metric field (RFMF) theory of non-retinotopic visual perception. *Perception, ECVP Abstract, 42*, 160.
Ogmen, H., Otto, T. U., & Herzog, M. H. (2006). Perceptual grouping induces non-retinotopic feature attribution in human vision. *Vision Research, 46*(19), 3234–3242.
Olson, C. R. (2003). Brain representation of object-centered space in monkeys and humans. *Annual Review of Neuroscience, 26*, 331–354.
Otto, T. U., Ogmen, H., & Herzog, M. H. (2010). Perceptual learning in a nonretinotopic frame of reference. *Psychological Science, 21*(8), 1058–1063.
Pack, C., Grossberg, S., & Mingolla, E. (2001). A neural model of smooth pursuit control and motion perception by cortical area MST. *Journal of Cognitive Neuroscience, 13*(1), 102–120.
Pola, J., & Wyatt, H. J. (1989). The perception of target motion during smooth pursuit eye movements in the open-loop condition: Characteristics of retinal and extraretinal signals. *Vision Research, 29*(4), 471–483.
Poljac, E., Verfaillie, K., & Wagemans, J. (2011). Integrating biological motion: The role of grouping in the perception of point-light actions. *PloS One, 6*(10), e25867.

Proffitt, D. R., Cutting, J. E., & Stier, D. M. (1979). Perception of wheel-generated motions. *Journal of Experimental Psychology. Human Perception and Performance, 5*(2), 289–302.
Rushton, S. K., Bradshaw, M. F., & Warren, P. A. (2007). The pop out of scene-relative object movement against retinal motion due to self-movement. *Cognition, 105*(1), 237–245.
Sereno, M. I., Dale, A. M., Reppas, J. B., Kwong, K. K., Belliveau, J. W., Brady, T. J., et al. (1995). Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science, 268*(5212), 889–893.
Shimozaki, S. S., Eckstein, M., & Thomas, J. P. (1999). The maintenance of apparent luminance of an object. *Journal of Experimental Psychology. Human Perception and Performance, 25*(5), 1433–1453.
Shum, K., & Wolford, G. (1983). A quantitative study of perceptual vector analysis. *Perception & Psychophysics, 34*(1), 17–24.
Snowden, R. J. (1992). Sensitivity to relative and absolute motion. *Perception, 21*(5), 563–568.
Sokolov, A., & Pavlova, M. (2006). Visual motion detection in hierarchical spatial frames of reference. *Experimental Brain Research, 174*(3), 477–486.
Souman, J. L., Hooge, I. T. C., & Wertheim, A. (2005). Vertical object motion during horizontal ocular pursuit: Compensation for eye movements increases with presentation duration. *Vision Research, 45*, 845–853.
Souman, J. L., Hooge, I. T. C., & Wertheim, A. H. (2006a). Frame of reference transformations in motion perception during smooth pursuit eye movements. *Journal of Computational Neuroscience, 20*, 61–76.
Souman, J. L., Hooge, I. T. C., & Wertheim, A. H. (2006b). Localization and motion perception during smooth pursuit eye movements. *Experimental Brain Research, 171*(4), 448–458.
Stoper, A. E. (1973). Apparent motion of stimuli presented stroboscopically during pursuit movement of the eye. *Perception & Psychophysics, 13*(2), 201–211.
Swanston, M., Wade, N., & Day, R. (1987). The representation of uniform motion in vision. *Perception, 16*(2), 143–159.
Tadin, D., Lappin, J. S., Blake, R., & Grossman, E. D. (2002). What constitutes an efficient reference frame for vision? *Nature Neuroscience, 5*, 1010–1015.
Takemura, H., Ashida, H., Amano, K., Kitaoka, A., & Murakami, I. (2012). Neural correlates of induced motion perception in the human brain. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience, 32*(41), 14344–14354.
Thurman, S., & Lu, H. (2013). Complex interactions between spatial, orientation, and motion cues for biological motion perception across visual space. *Journal of Vision, 13*, 1–18.
Tootell, R. B., Silverman, M., Switkes, E., & De Valois, R. (1982). Deoxyglucose analysis of retinotopic organization in primate striate cortex. *Science, 218*(4575), 902–904.
Turano, K. A., & Heidenreich, S. M. (1996). Speed discrimination of distal stimuli during smooth pursuit eye motion. *Vision Research, 36*(21), 3507–3517.
Turano, K. A., & Heidenreich, S. M. (1999). Eye movements affect the perceived speed of visual motion. *Vision Research, 39*(6), 1177–1187.
Turano, K. A., & Massof, R. W. (2001). Nonlinear contribution of eye velocity to motion perception. *Vision Research, 41*(3), 385–395.
Wade, N., & Swanston, M. (1987). The representation of nonuniform motion in vision. *Perception, 16*(5), 555–571.
Wallach, H. (1959). The perception of motion. *Scientific American, 201*(1), 56–60.
Wallach, H., O'Leary, A., & McMahon, M. L. (1982). Three stimuli for visual motion perception compared. *Perception & Psychophysics, 32*(1), 1–6.
Warren, P. A., & Rushton, S. K. (2009). Optic flow processing for the assessment of object movement during ego movement. *Current Biology: CB, 19*(18), 1555–1560.
Wertheim, A. (1987). Retinal and extraretinal information in movement perception: How to invert the Filehne illusion. *Perception, 16*(3), 299–308.
Wertheim, A. (1994). Motion perception during selfmotion: The direct versus inferential controversy revisited. *Behavioral and Brain Sciences, 17*(2), 293–355.
Yin, C., Shimojo, S., Moore, C., & Engel, S. A. (2002). Dynamic shape integration in extrastriate cortex. *Current Biology: CB, 12*(16), 1379–1385.
Zaehle, T., Jordan, K., Wüstenberg, T., Baudewig, J., Dechent, P., & Mast, F. W. (2007). The neural basis of the egocentric and allocentric spatial frame of reference. *Brain Research, 1137*(1), 92–103.